APPLICATION OF THE ADJOINT SYSTEM OF DIFFERENTIAL EQUATIONS IN THE SOLUTION OF THE BANG - BANG CONTROL PROBLEM

THOMAS RICHARD MC CALLA

APPLICATION OF THE ADJOINT SYSTEM OF DIFFERENTIAL EQUATIONS IN THE SOLUTION OF THE BANG-BANG CONTROL PROBLEM

Thomas Richard McCalla

APPLICATION OF THE ADJOINT SYSTEM OF DIFFERENTIAL EQUATIONS IN THE SOLUTION OF THE BANG-BANG CONTROL PROBLEM

by

Thomas Richard McCalla

Lieutenant, United States Naval Reserve

Submitted in partial fulfillment of
the requirements for the degree of

MASTER OF SCIENCE
IN
MATHEMATICS

United States Naval Postgraduate School

Monterey, California

1 9 6 1

APPLICATION OF THE ADJOINT SYSTEM OF DIFFERENTIAL EQUATIONS IN THE SOLUTION OF THE BANG-BANG CONTROL PROBLEM

by

Thomas Richard McCalla

This work is accepted as fulfilling
the thesis requirements for the degree of
MASTER OF SCIENCE
IN
MATHEMATICS
from the
United States Naval Postgraduate School

ABSTRACT

Some problems in the optimum control of a linear dynamic system[1] are investigated, particularly the problem of determining the minimum time required to drive a linear, constant-coefficient dynamic system from an initial state to a specified terminal state with a limited power source. The important feature of the paper is that an elementary method of solution for this problem is given. It is a method of successive approximations based on the adjoint system of differential equations in a way similar to that which Bliss used in calculating differentials in Ballistics. A program is given for solving the minimum-time problem on a digital computer; an elementary proof is given that if the routine converges then the solution thus found yields the desired minimum time.

The forcing vector is assumed to have bounded components, representing limited driving power or voltage. This forcing vector is to be selected as a function of time so as to drive the dynamic system from one state to another in minimum time. In problems where certain control variables are bounded, it is well known that the control function assumes at all times either its maximum or its minimum value, changing discontinuously from one to the other. This is generally the case here. The problems with this type of control are frequently referred to as Bang-Bang Control problems [1],[2],[3][2].

1. A system of differential equations will be considered as representing a dynamic system.

2. References to the bibliography will be denoted by numbers in brackets [ ].

The solution of the minimum-time problem for an N-th order linear, constant-coefficient system, in which the matrix of coefficients has real and distinct eigenvalues, is given in matrix notation, including the set of differential-correction equations for determining the adjoint vector and minimum time associated with the desired terminal state. The solution of the adjoint system is given in terms of exponential functions of the eigenvalues of the matrix of coefficients and the corresponding eigenvectors (see Appendix I), the components of which are in the same ratio as the cofactors of the elements of a column of the corresponding characteristic matrix of the original system (see Appendix II).

The general case of the second-order minimum-time problem was programmed and run on the CDC-1604 computer. Because the convergence of the differential-correction equations may be critical, especially for stable systems, a search routine was incorporated into the program to search the input-parameter space for "useable" starting values. A second supplementary program provides for the plotting of minimum-time trajectories[1] with initial state $X(0) = 0$ for any second-order system.

The general third-order minimum-time problem was also programmed and run on the CDC-1604, using the general matrix techniques described for the N-th order system. Extension of the program to N-th order could be accomplished directly.

1. A minimum-time trajectory is the locus of points in the state space through which the dynamic system passes as it is driven from an initial state to a terminal state in minimum time.

A plot of the minimum-time trajectories with initial state $X(0) = 0$ for a particular stable system is included in illustration 2. This illustration also contains the switching curves[1] associated with this initial state and the region of static stability.[2] Switching curves do not seem to have a very general significance, i.e., they are not implicit in the differential equations but depend on the initial conditions. It is to be noted however that for a second-order dynamic system and those minimum-time trajectories with two switching times, the time between the switching times is an invariant of the dynamic system.

The method of differential corrections for the adjoint system was first developed by Bliss for making differential corrections in Ballistics [8]. The set of differential-correction equations developed in this paper provide for the correcting of the input parameters to obtain the adjoint vector and minimum time associated with specified initial and terminal states. None of the papers reviewed on the Bang-Bang Control Problem provide such a system.

I would like to express my gratitude for the encouragement, guidance, and inspiration which Professor Faulkner has provided, and without which this paper would not have been possible.

1. A switching curve (surface) is the locus of points at which one of the components of the optimum forcing function changes sign.

2. A point of static stability of the system represented by $\dot{X} = AX + F$ is a point such that an admissible force can be chosen to constrain the system to that point. That is, it is a point where $|\sum_j a_{ij}x_j| \leq 1$, $(i=1,2,\ldots,n)$, and $\dot{x}_i = 0$, $(i=1,2,\ldots,n)$ for some admissible force F (see section 4).

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

1. Introduction.

Let us consider a linear dynamic system represented by the system of differential equations in matrix form

$$(1) \quad \dot{X}(t) = A\,X(t) + F(t)$$

or componentwise as

$$(2) \quad \dot{x}_i(t) = \sum_k a_{ik}x_k(t) + f_i(t) \qquad (i = 1,2,\ldots,n).$$

The matrix A is an n x n matrix with constant elements a_{ik} and with real, distinct eigenvalues.

X(t) is an n x 1 matrix representing a point or a <u>state</u> of the dynamic system. No distinction will be made generally between an n x 1 (column) matrix and a vector. If a quantity is to be considered as a vector, as in a scalar product, a bar (—) will be placed over it.

The dynamic system is controlled by a limited power source, i.e., the components $f_i(t)$ of the vector forcing (control) function F(t) are restricted by the bounds

$$(3) \quad |f_i(t)| \leq L_i \qquad (i = 1,2,\ldots,n)$$

where the L_i are preassigned constants. The L_i may all be assumed to be 1 (one) without loss of generality, and this will be the case henceforth.

The principal problem under investigation is the determination of the minimum time required to drive the dynamic system from an initial state X(0) to a specified terminal state X_f, and the determination of the control force F(t) which achieves this minimum time and the corresponding path or trajectory.

The solution is effected by using the adjoint system of differential equations. Certain properties of the adjoint system make this method well suited for the solution of this problem.

The adjoint system of differential equations is a homogeneous system; the system adjoint to (1) and (2) is formally defined by Bliss [9] as

$$(4) \quad \dot{U}(t) = - A^T U(t)$$

where the matrix A need not have constant elements. The superscript T denotes the transpose of a matrix.

Written componentwise, the system adjoint to (1) and (2) is

$$(5) \quad \dot{u}_i(t) = - \sum_k a_{ki} u_k(t) \qquad (i = 1,2,\ldots,n).$$

Whenever the elements of A are constant the homogeneous adjoint system can be solved directly in terms of exponential functions of the eigenvalues of the matrix of coefficients, $-A^T$, and the corresponding eigenvectors.[1]

The fundamental adjoint formula [9] of Bliss[2]

$$(6) \quad \left[\sum_i u_i x_i \right]_{t_1}^{t_2} = \int_{t_1}^{t_2} \left(\sum_i u_i f_i \right) dt$$

allows us, by proper selection of the adjoint solutions (choosing those solutions of the adjoint system which meet specified end conditions), to express the terminal values of the solutions of the original set of inhomogeneous equations in terms of integrals of the adjoint solutions and the forcing function of the dynamic system.

1. See Appendix I.
2. See section 2 for derivation and proof.

The order of development in this paper will be as follows:

First, the adjoint system of differential equations is defined, and the relation is established between solutions of the original system of equations and its adjoint through the fundamental adjoint formula, a form of Green's Formula[11].

Second, the minimum-time problem is solved by applying the forcing vector such that its projection on a properly-chosen adjoint vector is at all times a maximum. This property is also known as the Maximum Principle of Pontryagin. The solution requires the solution of a two-point boundary-value problem. The method employed is a method of successive approximations, using the differential procedures which Bliss introduced in Ballistics, based on the adjoint system [8]. Further, an elementary proof is given that a solution so obtained furnishes the minimum time.

Third, an example is given of the solution of a specific second-order system through the use of the adjoint system of differential equations.

Fourth, the solution of the general N-th order problem is derived in matrix notation for the case of a linear, constant-coefficient system of differential equations where the matrix of coefficients has real and distinct eigenvalues. This includes the differential-correction equations for determining the adjoint vector and the minimum time associated with a specified terminal state.

Fifth, the techniques described above are applied in the programming on a digital computer of the solution of the minimum-time problem for general second- and third-order systems.

2. The Adjoint System of Differential Equations, Green's Formula.

Let us consider a system of linear differential equations written in matrix form as

$$(1) \quad \dot{X} = A X + F$$

or componentwise as

$$(2) \quad \dot{x}_i = \sum_k a_{ik} x_k + f_i \qquad (i = 1,2,\ldots,n)$$

where the elements a_{ik} of matrix A are constants (or continuous functions of t), and where f_i are functions of t [9].

<u>Definition</u>:

The system of differential equations adjoint to (1) and (2) is defined [9] as

$$(3) \quad \dot{U} = - A^T U$$

or componentwise as

$$(4) \quad \dot{u}_i = - \sum_k a_{ki} u_k \qquad (i = 1,2,\ldots,n).$$

<u>Theorem (Fundamental Adjoint Formula)</u> [9]:

If
$$\begin{cases} (2) \quad \dot{x}_i = \sum_k a_{ik} x_k + f_i & (i = 1,2,\ldots,n) \\ (4) \quad \dot{u}_i = - \sum_k a_{ki} u_k & (i = 1,2,\ldots,n) \end{cases}$$

where (4) is the adjoint of (2),

then
$$(5) \quad \frac{d}{dt} \left(\sum_i u_i x_i \right) = \sum_i u_i f_i \quad ;$$

this must hold for every solution of (2) and (4).

<u>Proof</u>:

Consider the Lagrange Identity [11]

$$(6) \quad \frac{d}{dt} \left(\sum_i u_i x_i \right) = \sum_i \left(u_i \dot{x}_i + \dot{u}_i x_i \right)$$

Substituting (2) and (4) into (6) we get

$$(7)\quad \frac{d}{dt}\Big(\sum_i u_i x_i \Big) = \sum_i u_i\Big(\sum_k a_{ik} x_k + f_i \Big) + \sum_i x_i \Big(- \sum_k a_{ki} u_k \Big).$$

Expanding (7) we get

$$(8)\quad \frac{d}{dt}\Big(\sum_i u_i x_i \Big) = \sum_i \sum_k a_{ik} u_i x_k + \sum_i u_i f_i - \sum_i \sum_k a_{ki} u_k x_i \ ,$$

and, interchanging the meaning of i and k in the last term,

$$(9)\quad \frac{d}{dt}\Big(\sum_i u_i x_i \Big) = \sum_i \sum_k a_{ik} u_i x_k + \sum_i u_i f_i - \sum_k \sum_i a_{ik} u_i x_k \ ,$$

which upon interchanging the order of summation in the last term becomes

$$(10)\quad \frac{d}{dt}\Big(\sum_i u_i x_i \Big) = \sum_i \sum_k a_{ik} u_i x_k + \sum_i u_i f_i - \sum_i \sum_k a_{ik} u_i x_k \ .$$

Hence we get the final result, the Fundamental Adjoint Formula of Bliss [9]:

$$(5)\quad \frac{d}{dt}\Big(\sum_i u_i x_i \Big) = \sum_i u_i f_i \ .$$

Alternatively, the Fundamental Adjoint Formula can be written in integral form as

$$(11)\quad \Big[\sum_i u_i x_i \Big]_{t_1}^{t_2} = \int_{t_1}^{t_2} \Big(\sum_i u_i f_i \Big)\, dt \ .$$

This form is often referred to as Green's Formula [11].

3. Outline of Solution, Preliminary Formulas.

Let us now consider the system of differential equations discussed in section 1

(1) $\dot{X}(t) = A X(t) + F(t)$.

In this paper the dimension of F(t) is the same as that of X(t), though this is irrelevant to the method of solution.

Given an initial state X(0) of the dynamic system and another state X_f (referred to as the terminal state), we desire to determine how to drive the system from X(0) to X_f in minimum time by proper application of the control force F(t).

The solution of the minimum-time problem consists basically of the following steps:

a. Equations are obtained from Green's Formula[1] (2.11) to express the terminal values of the system state variables in terms of the initial state values, the time, and an integral involving the forcing function F(t) and the solutions of the adjoint system.

b. It is established next that if a solution is obtained which maximizes a certain integral and satisfies the desired end conditions it is the desired solution.

c. Only solutions are sought which maximize an integral of this form. However, there are some parameters in the integral, and these must be determined so as to yield the

1. Equations from other sections will be denoted by decimal numbers; the number preceding the decimal point indicates the section, and the number following indicates the equation number in that section.

desired end conditions. The problem reduces to the solution of a two-point boundary-value problem.

d. This problem is solved by a method of successive approximations. The corrections to the input parameters are obtained by a method based on Bliss' method of computing differentials in Ballistics.

DeSoer [2] proves the existence of a minimum-time path for stable systems.

The solution of the adjoint system obtained in the process defines the functional for the Maximum Principle of Pontryagin [7]. The integrand of the functional is the scalar product of the forcing vector and a vector defined by the solution of the adjoint system of differential equations.

We therefore have under investigation the dynamic system governed by the system of differential equations

(1) $\dot{X} = A X + F$.

The system adjoint to (1) can be written directly as

(2) $\dot{U} = - A^T U$.

From the Fundamental Adjoint Formula (Green's Formula) we get, choosing $t_1 = 0$, $t_2 = T$,

(3) $$\left[\sum_i u_i x_i \right]_0^T = \int_0^T \left(\sum_i u_i f_i \right) dt \ .$$

Let us now examine the solutions of the homogeneous adjoint system. From Appendix I we see that (for n distinct eigenvalues $-m_j$ of matrix $-A^T$) there are n distinct solutions U^j of the adjoint system of the form

(4) $U^j = e^{-m_j t} C^j$ $(j = 1,2,\ldots,n)$

where C^j is the eigenvector $[c_{1j}\ c_{2j}\ \ldots\ c_{nj}]^T$ associated with eigenvalue $-m_j$. $U^j = [u_1^j\ u_2^j\ \ldots\ u_n^j\ ]^T$.

Since the solutions U^j are independent solutions of the adjoint system, it follows that the general solution of the adjoint system can be written as a linear combination of the U^j, i.e.,

(5) $\quad U = \sum_j k_j U^j \quad$ where the k_j are arbitrary constants.

Substituting (4) into (5) results in

(6) $\quad U = \sum_j k_j e^{-m_j t} C^j.$

Suppose we now select a set of "special" adjoint solutions $U_j = [u_{1j}\ u_{2j}\ \ldots\ u_{nj}]^T$, $\qquad (j = 1,2,\ldots,n)$

which meet the following end conditions:

(7) $\quad u_{ij}(T) = \delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$.

We see then that the terminal values $x_j(T)$ of the solutions of the original system of equations can be determined by successively substituting the "special" solutions of the adjoint system into the fundamental formula (3). That is, substituting in U_j in (3), we get

(8) $\quad \left[\sum_i u_{ij} x_i \right]_0^T = \int_0^T \left(\sum_i u_{ij} f_i \right) dt$

which, since $u_{jj}(T) = 1$, $u_{ij}(T) = 0$, $i \neq j$, reduces to

(9) $\quad x_j(T) = \sum_i u_{ij}(0) x_i(0) + \int_0^T \left(\sum_i u_{ij} f_i \right) dt, \quad (j=1,2,\ldots,n).$

If we consider U_j as a vector in a cartesian coordinate set with base vectors $\bar{e}_i$, then

(10) $\bar{U}_j = \sum_i u_{ij}\bar{e}_i.$

If we also define a vector $\bar{X}$, the system state vector, by the equation

(11) $\bar{X} = \sum_i x_i\bar{e}_i$

then equation (9) can be written as

(12) $x_j(T) = (\bar{U}_j.\bar{X})_o + \int_o^T (\bar{U}_j.\bar{F})\,dt \qquad (j = 1,2,\ldots,n).$

Since the U_j are independent solutions of the adjoint system a linear combination of the U_j is also a solution,

(13) $U = U_1 + b_1U_2 + \ldots\ldots + b_{n-1}U_n$

where the b_j are arbitrary constants.

A solution U defined in this way is called an <u>adjoint vector</u>.

From (12) we have the following equations, obtained by successively substituting in the $\bar{U}_j$ into (3):

$$
(14)\begin{cases}
x_1(T) = (\bar{U}_1.\bar{X})_o + \int_o^T (\bar{U}_1.\bar{F})\,dt \\
x_2(T) = (\bar{U}_2.\bar{X})_o + \int_o^T (\bar{U}_2.\bar{F})\,dt \\
\vdots \\
x_n(T) = (\bar{U}_n.\bar{X})_o + \int_o^T (\bar{U}_n.\bar{F})\,dt
\end{cases}
$$

If we form any linear combination of the above, using say the b_j of (13), we get

$$(15) \quad x_1(T) + b_1 x_2(T) + \ldots + b_{n-1} x_n(T) = (\bar{U}.\bar{X})_o + \int_o^T (\bar{U}.\bar{F})dt$$

Equations (14) and (15) are the fundamental equations for the solution of the problem.

We shall hereafter consider only forcing vectors $\bar{F}$ such that the scalar product $(\bar{U}.\bar{F})$ is a maximum. In the next sections it will be seen how this characterizes the solutions.

4. A Theorem on Minimum Time.

In this section it will be shown that if an admissible trajectory can be found which satisfies a maximizing principle, then, at least under the conditions given below, it furnishes the desired minimum time.

An admissible trajectory is one which has the specified initial and terminal points; with it must be associated an admissible forcing function. An admissible forcing function **F** is one which satisfies the constraints on its components, $|f_i| \leq 1$. In this section no specific assumptions are made regarding the eigenvalues of the matrix A in the system of differential equations

$$\text{(1)} \qquad \dot{X} = A X + F$$

but A is assumed to be a constant matrix.

A point of static stability of (1) is a point such that an admissible force can be chosen to constrain the system to the point. That is, it is a point where $|\sum_j a_{ij}x_j| \leq 1$, $(i = 1,2,\ldots,n)$, and $\dot{x}_i = 0$, $(i = 1,2,\ldots,n)$ for some admissible F. A point such that $|\sum_j a_{ij}x_j| = 1$, $(i = 1,2,\ldots,n)$ is a vertex point of the points of static stability.

Now suppose we have in some way found a path C^+ and a vector U^+ which together satisfy the following relations:

H.1. The path C^+ is admissible for some value of T, say T^+.

H.2. $U^+ = U_1 + b_1U_2 + \ldots + b_{n-1}U_n$ is a solution to the adjoint system of differential equations. The U_j are the "special" solutions defined earlier, and the b's are known constants.

H.3. For each value of t, the force F^+ associated with C^+ is such that the scalar product $(\bar{F}^+ . \bar{U}^+)$ is a maximum.

H.4. Each component u_i^+ $(i = 1,2,\ldots,n)$ of U^+ vanishes at most at isolated points in $(0,T^+)$.

Theorem. The trajectory C^+ which satisfies H.1, H.2, H.3, and H.4 yields the minimum time T^+ and is unique, compared with all paths which contain a point of static stability that is not a vertex point.

Proof. Consider equation (3.15) with the values $b_1,\ldots,b_{n-1}$ from H.2 for any admissible trajectory for time T^+.

$$(2) \qquad X_{1f} + b_1 X_{2f} + \ldots + b_{n-1} X_{nf} - (\bar{U}^+ . \bar{X})_o = \int_o^{T^+} (\bar{U}^+ . \bar{F}^+) dt.$$

We shall see that F^+ maximizes this and then as a consequence T^+ is minimal. Note that every term on the left side of (2) is determined by the specified conditions X_o, X_f and the quantities $b_1,\ldots,b_{n-1}$, T^+.

Now consider the force F^+ above and any other force F' which is admissible with time T^+. Equation (2) must hold for

admissible path for that value T^+ of T.

There might conceivably be, however, a path C" with a smaller value T_e of T. Assume this to be the case and assume that there is a point X_s of static stability on C". Let $T^+ - T_e = \Delta T$, and let t_s be the time associated with X_s on C".

The force function F' constructed as follows then will lead to an admissible trajectory C' with $T = T^+$:

$$f_i' = \begin{cases} f_i''(t) & 0 < t < t_s \\ -\sum_j a_{ij} x_{js} & t_s < t < t_s + \Delta T \\ f_i''(t - \Delta T) & t_s + \Delta T < t < T^+. \end{cases}$$

The middle segment of C' constrains X to X_s for an interval of length ΔT, and the two end segments then lead from X_0 to X_s and from X_s to X_f.

Now note that

$$(4) \quad \int_0^{T^+} [\bar{U}^+ . (\bar{F}^+ - \bar{F}')]\, dt \geq \int_{t_s}^{t_s+\Delta T} [\bar{U}^+ . (\bar{F}^+ - \bar{F}')]\, dt,$$

and, for at least one value of i,

$$(5) \quad \int_{t_s}^{t_s+\Delta T} (f_i^+ - f_i')\, u_i^+ > 0$$

since u_i^+ has only isolated zeros and $(f_i^+ - f_i')u_i^+ = (1 \pm f_i')|u_i^+|$.

But C^+ is the only admissible path associated with time T^+. There is then a contradiction. It arises from the assumption that a second path with $T < T^+$ existed. This completes the proof: if such a path is found, it defines the solution and is unique.

Clearly the proof holds if in H.3, $(\overline{F}^+ . \overline{U}^+)$ is a minimum rather than a maximum on C^+.

Corollary. If the matrix A has real distinct eigenvalues, if either X_0 or X_f is a point of static stability which is not a vertex point, and if a curve C^+ and an adjoint vector have been found which satisfy H.1, H.2, H.3 with none of the b's of H.2 zero, then C^+ defines the minimum value of T and is unique.

In this case every trajectory contains a point of static stability X_0 or X_f which is not a vertex point. Every u_i^+ has the form $u_i^+ = \sum_j c_{ij}^+ e^{-m_j t}$ with at least one c_{ij}^+ distinct from zero for each value of i. It has at most n-1 zeros, and U^+ satisfies H.4. All of the hypotheses of the theorem are satisfied by C^+.

If the values of $b_1, \ldots, b_{n-1}$, T^+ were known, it would be just a matter of integration to get the trajectory by setting $f_i = \operatorname{sgn} u_i^+$. In the next sections these will be determined by a method of successive approximations.

The proof of the theorem in this section is due to Professor Faulkner.

5. Example of a Specific Second-Order Problem and Solution.

In this section the solution of the minimum-time problem for a specific second-order stable system will be worked out in full detail. The solution consists basically of the following steps:

a. The adjoint system of differential equations is solved.

b. A set of input parameters (T,b) is selected, and a set of "special" adjoint solutions (with the corresponding adjoint vector) is determined.

c. The terminal values x(T), y(T) of the dynamic system are expressed as integrals involving the "special" adjoint solutions and the vector forcing function, and a trajectory is determined by the Maximum Principle.

d. The resulting terminal state (x(T),y(T)) is compared with the required terminal state (X_f,Y_f), and the changes in T and b required to effect (X_f,Y_f) are determined from a set of differential-correction equations.

The dynamic system which will be considered is governed by the pair of linear first-order differential equations

$$(1) \quad \begin{cases} \dot{x} = -x + y + f & |f| \leq 1 \\ \dot{y} = -2x - 4y + g & |g| \leq 1. \end{cases}$$

The system of differential equations adjoint to (1) can be written directly as

$$(2) \quad \begin{cases} \dot{u} = u + 2v \\ \dot{v} = -u + 4v \end{cases}$$

[illegible]

5. The terminal values $x(t_f)$, $y(t_f)$ [illegible] are minimized in accordance with the "[illegible]" [illegible] solutions—the [illegible] is determined by the [illegible] principle.

6. The resulting terminal value $x(t_f)$, $y(t_f)$ is compared with the required terminal state (x_f, y_f), and the changes in ξ and η required to effect (x_f, y_f) are determined from a set of differential-correction equations.

The dynamic system which will be considered is described by the pair of linear homogeneous differential equations

(1) [illegible] (1.1) [illegible] (1.2)

The system of differential equations implied by (1) can be written formally as

(2) [illegible]

The characteristic equation of matrix A is

$$(3) \qquad \begin{vmatrix} -1-m & 1 \\ -2 & -4-m \end{vmatrix} = 0$$

The eigenvalues of A, roots of (3), are $m = -2, -3$.
Now, if m_1, m_2 are eigenvalues of matrix A, then $-m_1, -m_2$ are eigenvalues of $-A^T$.

Hence e^{2t}, e^{3t} and the eigenvalues of $-A^T$ will define a fundamental set of solutions of the adjoint system (2). Denote these solutions by U^1, U^2 where

$$(4) \qquad U^1 = \begin{bmatrix} u^1 \\ v^1 \end{bmatrix} \qquad U^2 = \begin{bmatrix} u^2 \\ v^2 \end{bmatrix}$$

From Appendix I we know that for each distinct eigenvalue of $-A^T$, the corresponding solution of the adjoint system can be written in the form

$$(5) \qquad U^j = e^{-m_j t} C^j$$

where $-m_j$ is an eigenvalue of $-A^T$, and C^j is the corresponding eigenvector.

Since the eigenvalues are real and distinct, we know (see Appendix II) that the components of C^j are in the same ratio as the cofactors of the elements of a column of the characteristic matrix $[A - m_j I]$. We can therefore determine that

$$(6) \qquad C^1 = \begin{bmatrix} c_{11} \\ c_{21} \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix} \qquad C^2 = \begin{bmatrix} c_{12} \\ c_{22} \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

Then the solutions U^1, U^2 of the adjoint system are

$$(7) \qquad U^1 = e^{-m_1 t} C^1 = e^{2t} \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

and

$$(8)\quad U^2 = e^{-m_2 t} C^2 = e^{3t}\begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

Since U^1, U^2 comprise a fundamental set for the adjoint system (2), $U = k_1U^1 + k_2U^2$ is also a solution, i.e.,

$$(9)\quad U = k_1e^{2t}\begin{bmatrix} 2 \\ 1 \end{bmatrix} + k_2e^{3t}\begin{bmatrix} 2 \\ 2 \end{bmatrix} = \begin{bmatrix} 2 & 2 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} k_1e^{2t} \\ k_2e^{3t} \end{bmatrix}$$

where the k's are arbitrary constants.

We will denote the matrix of eigenvectors of $-A^T$ by C, where

$$(10)\quad C = [C^1\ C^2] = \begin{bmatrix} 2 & 2 \\ 1 & 2 \end{bmatrix}$$

Inverting C we get $C^{-1} = \begin{bmatrix} 1 & -1 \\ -\frac{1}{2} & 1 \end{bmatrix}$

Let us now return to the original set of differential equations (1) which we may write as

$$(11)\quad \begin{cases} \dot{x} + x - y - f = 0 \\ \dot{y} + 2x + 4y - g = 0 \end{cases}$$

Suppose we now multiply the first equation of (11) by u and the second by v, where u,v are required only to be continuous functions of t which have continuous first derivatives. Adding the two products and integrating the sum from 0 to T, we get [4]

$$(12)\quad \int_0^T [u(\dot{x} + x - y - f) + v(\dot{y} + 2x + 4y - g)]dt = 0.$$

Note that if x,y are solutions of the original set of equations, then u,v used in this way are called Lagrange multipliers.

Integrating (12) by parts we get

$$(13)\quad [ux + vy]_0^T + \int_0^T [x(-\dot{u} + u + 2v) + y(-\dot{v} - u + 4v)]dt$$

$$- \int_0^T [uf + vg]\, dt = 0.$$

If we now choose u,v as solutions of the adjoint system, we have

$$(14)\quad \begin{cases} -\dot{u} + u + 2v = 0 \\ -\dot{v} - u + 4v = 0 \end{cases}$$

If u,v are thus chosen, (13) reduces to

$$(15)\quad [ux + vy]_0^T = \int_0^T [uf + vg]dt$$

Note that (15) is a particular example of the Fundamental Adjoint Formula of section 2.

Suppose we now choose a set of "special" solutions of the adjoint system as follows:

$$(16)\quad U_1 = \begin{bmatrix} u_1 \\ v_1 \end{bmatrix} \qquad \text{such that} \quad \begin{matrix} u_1(T) = 1 \\ v_1(T) = 0 \end{matrix}$$

and

$$(17)\quad U_2 = \begin{bmatrix} u_2 \\ v_2 \end{bmatrix} \qquad \text{such that} \quad \begin{matrix} u_2(T) = 0 \\ v_2(T) = 1 \end{matrix}$$

Substituting U_1 into (9) we get, evaluating at $t = T$

$$(18)\quad U_1(T) = \begin{bmatrix} u_1(T) \\ v_1(T) \end{bmatrix} = \begin{bmatrix} 2 & 2 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} k_{11}e^{2T} \\ k_{21}e^{3T} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

Note: The second subscripts of k_{11}, k_{21} indicate that they correspond to the first solution U_1.

Multiplying (18) through by C^{-1} we get

$$(19) \quad \begin{bmatrix} k_{11}e^{2T} \\ k_{21}e^{3T} \end{bmatrix} = C^{-1}\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} c'_{11} \\ c'_{21} \end{bmatrix} = \begin{bmatrix} 1 \\ -1/2 \end{bmatrix}$$

where the c'_{ij} are elements of C^{-1}.

Therefore $k_{11} = e^{-2T}$

$$k_{21} = \frac{e^{-3T}}{-2}$$

Then

$$(20) \quad \begin{bmatrix} u_1(t) \\ v_1(t) \end{bmatrix} = \begin{bmatrix} 2 & 2 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} e^{2(t-T)} \\ \dfrac{e^{3(t-T)}}{-2} \end{bmatrix}$$

Similarly, substituting U_2 into (9), we determine that

$$(21) \quad k_{12} = -e^{-2T}$$

$$k_{22} = e^{-3T}$$

so that

$$(22) \quad \begin{bmatrix} u_2(t) \\ v_2(t) \end{bmatrix} = \begin{bmatrix} 2 & 2 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} -e^{2(t-T)} \\ e^{3(t-T)} \end{bmatrix}$$

Substituting U_1 into (15), we get

$$(23) \quad x(T) = u_1(0)x(0) + v_1(0)y(0) + \int_0^T [u_1 f + v_1 g]dt.$$

Substituting U_2 into (15), we get

$$(24) \quad y(T) = u_2(0)x(0) + v_2(0)y(0) + \int_0^T [u_2 f + v_2 g]dt.$$

Let us now consider the "special" adjoint solutions as vector functions

$$(25) \quad \bar{U}_1 = u_1\bar{e}_1 + v_1\bar{e}_2$$

$$\bar{U}_2 = u_2\bar{e}_1 + v_2\bar{e}_2$$

where $\bar{e}_1$, $\bar{e}_2$ are the base vectors for a cartesian coordinate set in two dimensions.

Since U_1 and U_2 are independent solutions of the adjoint system, a linear combination of these is also a solution, i.e.,

$$(27)\quad \begin{aligned} \bar{U} &= \bar{U}_1 + b\bar{U}_2 \\ &= (u_1 + bu_2)\bar{e}_1 + (v_1 + bv_2)\bar{e}_2 \\ &= u\,\bar{e}_1 + v\,\bar{e}_2 \end{aligned}$$

Combining (23) and (24) in the same linear combination, we obtain

$$(28)\quad x(T) + b\,y(T) = (u_1+bu_2)_o x(0) + (v_1+bv_2)_o y(0) + \int_o^T [(u_1+bu_2)f + (v_1+bv_2)g]dt.$$

which can be written as

$$(29)\quad x(T) + b\,y(T) = u(0)x(0) + v(0)y(0) + \int_o^T [uf + vg]dt.$$

Defining vector $\bar{R} = x\bar{e}_1 + y\bar{e}_2$, we can rewrite (29) as

$$(30)\quad x(T) + b\,y(T) = (\bar{U}.\bar{R})_o + \int_o^T (\bar{U}.\bar{F})dt.$$

where $\bar{\mathbf{F}} = f\,\bar{e}_1 + g\,\bar{e}_2$.

Suppose now that we desire to drive the system from $(x(0),y(0))$ to (X_f,Y_f) in minimum time. As we have seen earlier (section 4), this may be accomplished by applying Pontryagin's Maximum Principle to (30), i.e., by choosing the vector $\bar{\mathbf{F}} = \bar{\mathbf{F}}^+$ such that the projection of $\bar{\mathbf{F}}^+$ on the adjoint vector $\bar{U}$ is at all times a maximum. That is, we choose $\bar{\mathbf{F}}^+$ such that its components satisfy the following relations:

$$(31)\quad \begin{cases} f^+ = \operatorname{sgn} u \\ g^+ = \operatorname{sgn} v. \end{cases}$$

Equation (30) then can be rewritten as

$$(32) \quad x(T) + b\, y(T) = (\bar{U}.\bar{R})_0 + \int_0^T (\bar{U}.\bar{F}^+)dt$$

while (23) can be rewritten as

$$(33) \quad x(T) = (\bar{U}_1.\bar{R})_0 + \int_0^T (\bar{U}_1.\bar{F}^+)dt ,$$

and (24) can be rewritten as

$$(34) \quad y(T) = (\bar{U}_2.\bar{R})_0 + \int_0^T (\bar{U}_2.\bar{F}^+)dt.$$

However, the value of T which we have selected is only a "guess" at the actual minimum time required to drive the system to (X_f,Y_f), and the constant b is only a "guess" for the linear combination of U_1,U_2 to determine the proper adjoint vector associated with the desired terminal state. We must therefore compare the resulting state $(x(T),y(T))$ with the desired terminal state (X_f,Y_f) to determine the errors

$$(35) \quad \begin{cases} X_f - x(T) = \Delta X \\ Y_f - y(T) = \Delta Y \end{cases}$$

and determine the changes δT, δb required in the input parameters to drive the system to the desired terminal state.

For this we will derive a set of equations to determine δT, δb. Suppose we replace T and b by nearby values

$$(36) \quad \begin{cases} T \longrightarrow T + \delta T \\ b \longrightarrow b + \delta b \end{cases}$$

Now, writing $x(T) = x(T,b)$ and $y(T) = y(T,b)$ since the terminal values are implicit functions of b, we get

$$(37)\quad \begin{cases} x(T,b) \longrightarrow x(T+\delta T, b+\delta b) = x(T,b) + \delta X \\ y(T,b) \longrightarrow y(T+\delta T, b+\delta b) = y(T,b) + \delta Y \end{cases}$$

plus terms of higher order.

Now

$$(38)\quad \begin{cases} \delta X \doteq \dfrac{\partial X}{\partial T}\delta T + \dfrac{\partial X}{\partial b}\delta b \\ \delta Y \doteq \dfrac{\partial Y}{\partial T}\delta T + \dfrac{\partial Y}{\partial b}\delta b \end{cases}$$

where δX, δY denote the total differentials of $x(T,b)$ and $y(T,b)$ respectively, i.e.,

$$(39)\quad \begin{cases} \dfrac{\partial}{\partial T}x(T) = \dfrac{\partial X}{\partial T} & \dfrac{\partial}{\partial b}x(T) = \dfrac{\partial X}{\partial b} \\ \dfrac{\partial}{\partial T}y(T) = \dfrac{\partial Y}{\partial T} & \dfrac{\partial}{\partial b}y(T) = \dfrac{\partial Y}{\partial b} \end{cases}$$

We can rewrite (38) as a single matrix equation

$$(40)\quad \begin{bmatrix} \delta X \\ \delta Y \end{bmatrix} \doteq \begin{bmatrix} \dfrac{\partial X}{\partial T} & \dfrac{\partial X}{\partial b} \\ \dfrac{\partial Y}{\partial T} & \dfrac{\partial Y}{\partial b} \end{bmatrix} \begin{bmatrix} \delta T \\ \delta b \end{bmatrix} \doteq \epsilon \begin{bmatrix} \delta T \\ \delta b \end{bmatrix}$$

where ϵ is the matrix defined in (40). This matrix is called the "correction" matrix.

Multiplying (40) through by ϵ^{-1} and setting $\delta X = \Delta X$ and $\delta Y = \Delta Y$, we get

$$(41)\quad \begin{bmatrix} \delta T \\ \delta b \end{bmatrix} \doteq \epsilon^{-1} \begin{bmatrix} \delta X \\ \delta Y \end{bmatrix} \doteq \epsilon^{-1} \begin{bmatrix} X_f - x(T) \\ Y_f - y(T) \end{bmatrix}$$

Equation (41) will be referred to as the "correction equation.

Hence we have determined the changes δT, δb in the input parameters required to drive the system to (X_f, Y_f), so that

$$(42) \quad \begin{cases} x(T + \delta T, b + \delta b) = X_f \\ y(T + \delta T, b + \delta b) = Y_f \end{cases}$$

except for higher-order terms.

Therefore we have solved the minimum-time problem assuming we can determine the elements of the "correction" matrix, and if the routine converges.

We will first determine $\frac{\partial X}{\partial T}$.

where $X = x(T) = (\bar{U}_1 . \bar{R})_o + \int_o^T (\bar{U}_1 . \bar{F}^+) dt.$

On a minimum-time trajectory each component of a two-dimensional forcing vector will switch (change sign) at most one time if the eigenvalues are real. These switching times are defined by the equations

$$(43) \quad \begin{cases} u(t_1) = 0, & f \text{ switches from } \pm 1 \text{ to } \mp 1 \text{ at } t = t_1. \\ v(t_2) = 0, & g \text{ switches from } \pm 1 \text{ to } \mp 1 \text{ at } t = t_2. \end{cases}$$

Order t_1, t_2 as t_k, $\quad 0 < k \leq K \leq 2$

Define $\begin{cases} 0 = t_0 & k = 0 \\ T = t_{K+1} & k = K \end{cases}$

so that $0 \leq k \leq K \leq 2$.

With the above notation, we can express (33) as

$$(44) \quad x(T) = (\bar{U}_1 . \bar{R})_o + \sum_{k=0}^{K} \int_{t_k}^{t_{k+1}} (\bar{U}_1 . \bar{F}^+) dt$$

where the integrand is continuous over each subinterval of integration t_k to t_{k+1}.

To simplify notation we will make the following definition:

$$(45)\qquad I_k = (\bar{U}_1 \cdot \bar{F}^+) , \qquad\qquad t_k < t < t_{k+1}$$

Then (44) can be rewritten as

$$(46)\qquad X = (\bar{U}_1 \cdot \bar{R})_o + \sum_{o}^{K} \int_{t_k}^{t_{k+1}} I_k \, dt.$$

Applying Leibniz' Rule for the differentiation of a definite integral the limits of which are functions of a parameter [5], we obtain

$$(47)\qquad \frac{\partial X}{\partial T} = \left(\frac{\partial \bar{U}_1}{\partial T} \cdot \bar{R}\right)_o + \sum_{o}^{K} \int_{t_k}^{t_{k+1}} \frac{\partial I_k}{\partial T} \, dt + \sum_{o}^{K} I_k(t_{k+1}) \frac{\partial t_{k+1}}{\partial T} - \sum_{o}^{K} I_k(t_k) \frac{\partial t_k}{\partial T}$$

or if $[\quad]_{t_k}$ denotes the discontinuity (jump) in a function at point t_k, we can write (47) as

$$(48)\qquad \frac{\partial X}{\partial T} = \left(\frac{\partial \bar{U}_1}{\partial T} \cdot \bar{R}\right)_o + \int_o^T \left(\frac{\partial \bar{U}_1}{\partial T} \cdot \bar{F}^+\right) dt + (\bar{U}_1 \cdot \bar{F}^+)_T - \sum_{1}^{K} (\bar{U}_1 \cdot [\bar{F}^+])_{t_k} \frac{\partial t_k}{\partial T}$$

where $K = 0$, 1, or 2.

Now, for fixed b, it is readily seen that $\frac{\partial t_k}{\partial T} = 1$, so so that (48) reduces to

$$(49)\qquad \frac{\partial X}{\partial T} = \left(\frac{\partial \bar{U}_1}{\partial T} \cdot \bar{R}\right)_o + \int_o^T \left(\frac{\partial \bar{U}_1}{\partial T} \cdot \bar{F}^+\right) dt + (\bar{U}_1 \cdot \bar{F}^+)_T - \sum_{1}^{K} (\bar{U}_1 \cdot [\bar{F}^+])_{t_k}$$

Similarly,

$$(50) \quad \frac{\partial Y}{\partial T} = \left(\frac{\partial U_2}{\partial T}.h\right)_o + \int_o^T \left(\frac{\partial U_2}{\partial T}.F^+\right)dt + (U_2.F^+)_T - \sum_1^h (U_2.[F^+])_{t_k}$$

By the same methods we can determine that $\frac{\partial X}{\partial b}$ and $\frac{\partial Y}{\partial b}$ can be expressed as

$$(51) \quad \frac{\partial X}{\partial b} = -\sum_{k=0}^{K-1} (U_1.[F^+])_{t_{k+1}} \frac{\partial t_{k+1}}{\partial b}$$

and

$$(52) \quad \frac{\partial Y}{\partial b} = -\sum_{k=0}^{K-1} (U_2.[F^+])_{t_{k+1}} \frac{\partial t_{k+1}}{\partial b}$$

To completely determine $\frac{\partial X}{\partial b}$ and $\frac{\partial Y}{\partial b}$ we must determine $\frac{\partial t_{k+1}}{\partial b}$

These can be determined by making the following linear approximation of the change in the switching times, denoted by δt_1 (and δt_2):

$$(53) \quad \delta t_1 = -\left.\frac{\delta u}{\dot{u}}\right|_{t_1} = \left[-\frac{1}{\dot{u}}\left(\frac{\partial u}{\partial T}\delta T + \frac{\partial u}{\partial b}\delta b\right)\right]_{t_1}$$

and since $u = u(T-t)$, it follows that $\dot{u} = -\frac{\partial u}{\partial T}$

so that (53) reduces to

$$(54) \quad \delta t_1 = \delta T - \left.\frac{\frac{\partial u}{\partial b}}{\dot{u}}\right|_{t_1} \delta b = \delta T - \left.\frac{u_2}{\dot{u}}\right|_{t_1} \delta b$$

so that

$$(55) \quad \frac{\partial t_1}{\partial b} = -\left.\frac{u_2}{\dot{u}}\right|_{t_1}$$

Similarly,

$$(56) \quad \frac{\partial t_2}{\partial b} = -\left.\frac{v_2}{\dot{v}}\right|_{t_2}$$

Having determined $\frac{\partial t_{k+1}}{\partial b}$ we can now write the final expressions of $\frac{\partial X}{\partial b}$ and $\frac{\partial Y}{\partial b}$ as

$$(57) \quad \frac{\partial X}{\partial b} = -\left\{ \left[u_1[f^+] \left(\frac{-u_2}{\dot{u}} \right) \right]_{t_1} + \left[v_1[g^+] \left(\frac{-v_2}{\dot{v}} \right) \right]_{t_2} \right\}$$

and

$$(58) \quad \frac{\partial Y}{\partial b} = -\left\{ \left[u_2[f^+] \left(\frac{-u_2}{\dot{u}} \right) \right]_{t_1} + \left[v_2[g^+] \left(\frac{-v_2}{\dot{v}} \right) \right]_{t_2} \right\}$$

We have now determined all the elements of the "correction" matrix, so that the minimum-time problem is solved, if the routine converges.

In this section we have developed the solution of the minimum-time problem for a specific second-order system. However, each step of the solution has been derived in general notation before substituting in the numerical values. It is therefore unnecessary to repeat the development for the general case for second-order systems. In the succeeding sections, the solution of the problem for the general case will be developed for higher-order systems.

An interesting property of second-order systems was discovered; it is the following: If a minimum-time trajectory for second-order systems has two switching times, the time between these switching times is an invariant of the dynamic system. An elementary proof of this result is given in section 8.

6. Derivation of Solution for Third-Order System.

Suppose we have a set of three inhomogeneous linear first-order differential equations with constant coefficients representing a dynamic system

(1) $\dot{X} = A X + F$

where $X = [x_1 \; x_2 \; x_3]^T$, $F = [f_1 \; f_2 \; f_3]^T$, and A is a 3 x 3 matrix with constant elements a_{ik}. The eigenvalues of A are assumed to be real and distinct.

The characteristic equation associated with matrix A is

(2) $| A - mI | = 0,$

i.e.,

(3) $$\begin{vmatrix} a_{11}-m & a_{12} & a_{13} \\ a_{21} & a_{22}-m & a_{23} \\ a_{31} & a_{32} & a_{33}-m \end{vmatrix} = 0 ,$$

which, when expanded by determinants, becomes a cubic equation in m of the form

(4) $m^3 + b_1 m^2 + b_2 m + b_3 = 0.$

Denote by m_1, m_2, m_3 the roots of (4), i.e., the eigenvalues of matrix A.

Then $-m_1$, $-m_2$, $-m_3$ are eigenvalues of $-A^T$.

Therefore, if we consider the system adjoint to (1), which is

(5) $\dot{U} = -A^T U$ where $U = [u_1 \; u_2 \; u_3]^T$

then $e^{-m_1 t}$, $e^{-m_2 t}$, $e^{-m_3 t}$ and the eigenvectors of $-A^T$ will define a fundamental set of solutions of the adjoint system.

Let us consider the solution of the adjoint system (5) in terms of the eigenvectors of the matrix of coefficients, $-A^T$. We have

(5) $\dot{U} = -A^T U.$

Assume a solution of the form

(6) $U = e^{rt} C'$ where $C' = [c_1 \; c_2 \; c_3]^T$ is a constant vector.

Differentiating (6) with respect to t, we obtain

(7) $\dot{U} = r e^{rt} C'$

Substitution of (7) and (6) back into (5) gives

(8) $r e^{rt} C' = -A^T e^{rt} C'$;

and since e^{rt} is a scalar quantity appearing on both sides of the equation, it follows that

(9) $r C' = -A^T C'.$

Therefore, if we choose r as an eigenvalue, $-m_j$, of $-A^T$ then C' is by definition the corresponding eigenvector. Hence, for each eigenvalue of $-A^T$ we have a unique solution of the adjoint system if the eigenvalues are distinct. Therefore, we have the following fundamental set of solutions of (5):

$$U^1 = e^{-m_1 t} C^1$$

$$U^2 = e^{-m_2 t} C^2$$

$$U^3 = e^{-m_3 t} C^3$$

Then any linear combination of U^1, U^2, U^3 is also a solution

(10) $U = k_1 e^{-m_1 t} C^1 + k_2 e^{-m_2 t} C^2 + k_3 e^{-m_3 t} C^3$

where the k's are arbitrary constants. If (10) is written in expanded form, we get

(11) $$U = k_1 e^{-m_1 t} \begin{bmatrix} c_{11} \\ c_{21} \\ c_{31} \end{bmatrix} + k_2 e^{-m_2 t} \begin{bmatrix} c_{12} \\ c_{22} \\ c_{32} \end{bmatrix} + k_3 e^{-m_3 t} \begin{bmatrix} c_{13} \\ c_{23} \\ c_{33} \end{bmatrix}$$

Equation (11) can then be expressed as the product of two matrices, as follows:

$$\begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{bmatrix} \begin{bmatrix} k_1 e^{-m_1 t} \\ k_2 e^{-m_2 t} \\ k_3 e^{-m_3 t} \end{bmatrix}$$

where
$$C = \begin{bmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{bmatrix}$$
is the matrix of eigenvectors of $-A^T$.[1]

Then, the general solution of the adjoint system can be written as

$$(12)\qquad \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} = C \begin{bmatrix} k_1 e^{-m_1 t} \\ k_2 e^{-m_2 t} \\ k_3 e^{-m_3 t} \end{bmatrix}$$

We will now choose a "special" solution $U_1 = [u_{11}\ u_{21}\ u_{31}]^T$ such that $u_{11}(T) = 1$, $u_{21}(T) = 0$, $u_{31}(T) = 0$. Substitution of U_1 into (12) evaluated at $t = T$ gives us

$$(13)\qquad \begin{bmatrix} u_{11}(T) \\ u_{21}(T) \\ u_{31}(T) \end{bmatrix} = C \begin{bmatrix} k_{11} e^{-m_1 T} \\ k_{21} e^{-m_2 T} \\ k_{31} e^{-m_3 T} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

where the second subscripts of the k's indicate constants corresponding to solution U_1.

Multiplying (13) through by C^{-1} we get

$$(14)\qquad \begin{bmatrix} k_{11} e^{-m_1 T} \\ k_{21} e^{-m_2 T} \\ k_{31} e^{-m_3 T} \end{bmatrix} = C^{-1} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} c'_{11} \\ c'_{21} \\ c'_{31} \end{bmatrix}$$

where the c'_{ij} are elements of C^{-1}.

1. See Appendix II for determination of components of C^j in terms of cofactors of $[A - m_j I]$.

Then

$$(15)\quad \begin{cases} k_{11} = c'_{11}\, e^{m_1 T} \\ k_{21} = c'_{21}\, e^{m_2 T} \\ k_{31} = c'_{31}\, e^{m_3 T} \end{cases}$$

Substitution of (15) back into (12) gives

$$(16)\quad \begin{bmatrix} u_{11} \\ u_{21} \\ u_{31} \end{bmatrix} = C \begin{bmatrix} k_{11} e^{-m_1 t} \\ k_{21} e^{-m_2 t} \\ k_{31} e^{-m_3 t} \end{bmatrix} = C \begin{bmatrix} c'_{11}\, e^{m_1 (T-t)} \\ c'_{21}\, e^{m_2 (T-t)} \\ c'_{31}\, e^{m_3 (T-t)} \end{bmatrix}$$

Similarly, we can choose a second "special" solution $U_2 = [u_{12}\ u_{22}\ u_{32}]^T$ of the adjoint system satisfying the end conditions $u_{12}(T) = 0$, $u_{22}(T) = 1$, $u_{32}(T) = 0$, which results in

$$(17)\quad \begin{bmatrix} u_{12} \\ u_{22} \\ u_{32} \end{bmatrix} = C \begin{bmatrix} c'_{12}\, e^{m_1 (T-t)} \\ c'_{22}\, e^{m_2 (T-t)} \\ c'_{32}\, e^{m_3 (T-t)} \end{bmatrix}$$

and, choosing a third "special" adjoint solution $U_3 = [u_{13}\ u_{23}\ u_{33}]^T$ satisfying the end conditions $u_{13}(T) = 0$, $u_{23}(T) = 0$, $u_{33}(T) = 1$, we obtain

$$(18)\quad \begin{bmatrix} u_{13} \\ u_{23} \\ u_{33} \end{bmatrix} = C \begin{bmatrix} c'_{13}\, e^{m_1 (T-t)} \\ c'_{23}\, e^{m_2 (T-t)} \\ c'_{33}\, e^{m_3 (T-t)} \end{bmatrix}$$

If we group (16), (17), and (18) into a single matrix equation, we obtain

$$(19)\quad \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ u_{21} & u_{22} & u_{23} \\ u_{31} & u_{32} & u_{33} \end{bmatrix} = C \begin{bmatrix} c'_{11} e^{m_1 (T-t)} & c'_{12} e^{m_1 (T-t)} & c'_{13} e^{m_1 (T-t)} \\ c'_{21} e^{m_2 (T-t)} & c'_{22} e^{m_2 (T-t)} & c'_{23} e^{m_2 (T-t)} \\ c'_{31} e^{m_3 (T-t)} & c'_{32} e^{m_3 (T-t)} & c'_{33} e^{m_3 (T-t)} \end{bmatrix}$$

$$(20)\quad [U] = C \begin{bmatrix} e^{m_1 (T-t)} & 0 & 0 \\ 0 & e^{m_2 (T-t)} & 0 \\ 0 & 0 & e^{m_3 (T-t)} \end{bmatrix} C^{-1}$$

$$(21)\quad [U] = C\ E\ C^{-1}$$

where E is the diagonal matrix with j^{th} diagonal element $e^{m_j (T-t)}$. and $[U] = [U_1\ U_2\ U_3]$.

Since U_1, U_2, U_3 are solutions of the adjoint system then any linear combination of these is also a solution:

$$(22) \quad U = U_1 + b_1 U_2 + b_2 U_3$$

where $U = [u_1 \ u_2 \ u_3]^T$;

$$\text{i.e.,} \quad \begin{cases} u_1 = u_{11} + b_1 u_{12} + b_2 u_{13} \\ u_2 = u_{21} + b_1 u_{22} + b_2 u_{23} \\ u_3 = u_{31} + b_1 u_{32} + b_2 u_{33}. \end{cases}$$

Written in matrix form, the above set of equations becomes

$$(23) \quad \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} = \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ u_{21} & u_{22} & u_{23} \\ u_{31} & u_{32} & u_{33} \end{bmatrix} \begin{bmatrix} 1 \\ b_1 \\ b_2 \end{bmatrix}.$$

Let $\bar{e}_1$, $\bar{e}_2$, $\bar{e}_3$ represent the base vectors of a three-dimensional cartesian coordinate system. We can then express U_1, U_2, U_3 as the vectors

$$(24) \quad \begin{aligned} \bar{U}_1 &= u_{11}\bar{e}_1 + u_{21}\bar{e}_2 + u_{31}\bar{e}_3 \\ \bar{U}_2 &= u_{12}\bar{e}_1 + u_{22}\bar{e}_2 + u_{32}\bar{e}_3 \\ \bar{U}_3 &= u_{13}\bar{e}_1 + u_{23}\bar{e}_2 + u_{33}\bar{e}_3 \end{aligned}$$

Then $U = U_1 + b_1 U_2 + b_2 U_3$ can be written as

$$(25) \quad \bar{U} = (u_{11} + b_1 u_{12} + b_2 u_{13})\bar{e}_1 + (u_{21} + b_1 u_{22} + b_2 u_{23})\bar{e}_2 + (u_{31} + b_1 u_{32} + b_2 u_{33})\bar{e}_3$$

which can also be written as

$$(26) \quad \bar{U} = u_1\bar{e}_1 + u_2\bar{e}_2 + u_3\bar{e}_3$$

A solution $\bar{U}$ obtained in this way is called an <u>adjoint vector</u>.

If in the Fundamental Adjoint Formula we let $t_1 = 0$ and $t_2 = T$, we have, from (2.11),

$$(27) \quad \Big[\sum_i u_i x_i \Big]_0^T = \int_0^T \Big(\sum_i u_i f_i \Big) dt.$$

If we desire to determine $x_j(T)$, we choose the adjoint solution $U = U_j$. This reduces (27) to

$$(28) \quad x_j(T) = \left(\sum_i u_{ij}x_i\right)_o + \int_o^T \left(\sum_i u_{ij}f_i\right)dt \quad (j=1,2,3),$$

which can be written in equivalent vector notation as

$$(29) \quad x_j(T) = (\bar{U}_j.\bar{X})_o + \int_o^T (\bar{U}_j.\bar{F})dt, \qquad (j=1,2,3),$$

where $\bar{X} = x_1\bar{e}_1 + x_2\bar{e}_2 + x_3\bar{e}_3$.

We wish to drive the system from state $X(0)$ to X_f in minimum time T. We therefore consider a linear combination of the $x_j(T)$, using say the constants b_j of (22). We get

$$(30) \quad x_1(T) + b_1x_2(T) + b_2x_3(T) = [(\bar{U}_1 + b_1\bar{U}_2 + b_2\bar{U}_3).\bar{X}]_o + \int_o^T [(\bar{U}_1 + b_1\bar{U}_2 + b_2\bar{U}_3).\bar{F}]dt,$$

which can be written as

$$(31) \quad x_1(T) + b_1x_2(T) + b_2x_3(T) = (\bar{U}.\bar{X})_o + \int_o^T (\bar{U}.\bar{F})dt.$$

Applying Pontryagin's Maximum Principle, i.e., by choosing $F = F^+$ such that $f_i^+ = \operatorname{sgn} u_i$, we obtain

$$(32) \quad x_1(T) + b_1x_2(T) + b_2x_3(T) = (\bar{U}.\bar{X})_o + \int_o^T (\bar{U}.\bar{F}^+)dt$$

We now compare the resulting terminal state $X(T)$ with the desired terminal state X_f, and define the component errors as

$$(33) \quad \begin{cases} X_{1f} - x_1(T) = \Delta X_1 \\ X_{2f} - x_2(T) = \Delta X_2 \\ X_{3f} - x_3(T) = \Delta X_3 . \end{cases}$$

If any ΔX_j $(j = 1,2,3)$ is not zero, we must determine the changes δT, δb_1, δb_2 in the input parameters required to effect X_f. Suppose the initial input parameters are replaced by nearby values

$$(34)\quad \begin{cases} T \longrightarrow T + \delta T \\ b_1 \longrightarrow b_1 + \delta b_1 \\ b_2 \longrightarrow b_2 + \delta b_2 . \end{cases}$$

Then the resulting changes δX_j in the terminal values $x_j(T)$ are determined to be (for $j = 1,2,3$)

$$(35)\quad \delta X_j = \left(\left(\frac{\partial \bar{U}}{\partial T}j \cdot \bar{X}\right)_o + \int_o^T \left(\frac{\partial \bar{U}}{\partial T}j \cdot \bar{F}^+\right) dt - \sum_{i=1}^{3} \sum_{k=1}^{K_i} (u_{ij}[f_i^+])_{t_{ik}}^{1} + f_j^+(T) \right) \delta T$$

$$+ \sum_{i=1}^{3} \sum_{k=1}^{K_i} \left(\frac{u_{ij}[f_i^+]\, u_{i2}}{\mathring{u}_i} \right)_{t_{ik}} \delta b_1$$

$$+ \sum_{i=1}^{3} \sum_{k=1}^{K_i} \left(\frac{u_{ij}[f_i^+]\, u_{i3}}{\dot{u}_i} \right)_{t_{ik}} \delta b_2$$

where K_i denotes the number of zeros of u_i, and t_{ik} the corresponding times. Equation (35) has the form

$$(36)\quad \delta X_j = \frac{\partial X}{\partial T}j\, \delta T + \frac{\partial X}{\partial b_1}j\, \delta b_1 + \frac{\partial X}{\partial b_2}j\, \delta b_2 \qquad (j=1,2,3)$$

Setting $\delta X_j = \Delta X_j$, and inverting the "correction" matrix defined by (36), we can determine the required changes δT, δb_1, δb_2 to effect X_f.

1. $(u_{ij}[f_i^+])_{t_{ik}} = u_{ij} f_i^+ \Big|_{t_{ik}+0} - u_{ij} f_i^+ \Big|_{t_{ik}-0}$

7. Derivation of Solution for N-th Order Systems.

Let us consider a linear dynamic system of the n^{th} order represented by the system of differential equations

$$(1) \qquad \overset{\circ}{X} = A X + F$$

where A is an n x n matrix with constant elements a_{ik} and with real and distinct eigenvalues. $X = [x_1\ x_2\ \ldots\ x_n]^T$ and $F = [f_1\ f_2\ \ldots\ f_n]^T$.

The system adjoint to (1) can be written directly as

$$(2) \qquad \overset{\circ}{U} = -A^T U$$

where $U = [u_1\ u_2\ \ldots\ u_n]^T$. We know from Appendix I that equation (2) has solutions of the form

$$(3) \qquad U = \sum_j k_j\, e^{-m_j t}\, C^j$$

where $-m_j$ are eigenvalues of matrix $-A^T$, and the C^j are the corresponding eigenvectors.

Equation (3) can also be written in matrix form as

$$(4) \qquad U = C \begin{bmatrix} k_1 e^{-m_1 t} \\ k_2 e^{-m_2 t} \\ \vdots \\ k_n e^{-m_n t} \end{bmatrix}$$

where C is the n x n matrix composed of the eigenvectors C^j, i.e., $C = [C^1\ C^2\ \ldots\ C^n]$ where the eigenvectors C^j are n x 1 column matrices.

If we choose a "special" solution $U_1 = [u_{11}\ u_{21}\ \ldots u_{n1}]^T$ to the adjoint such that $u_{11}(T) = 1$, $u_{21}(T) = \ldots = u_{n1}(T) = 0$, then equation (4) evaluated at t = T becomes

$$(5) \qquad \begin{bmatrix} u_{11}(T) \\ u_{21}(T) \\ \\ u_{n1}(T) \end{bmatrix} = C \begin{bmatrix} k_{11}e^{-m_1T} \\ k_{21}e^{-m_2T} \\ \\ k_{n1}e^{-m_nT} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ \\ 0 \end{bmatrix}$$

Multiplying (5) through by C^{-1} we get

$$(6) \qquad \begin{bmatrix} k_{11}e^{-m_1T} \\ k_{21}e^{-m_2T} \\ \\ k_{n1}e^{-m_nT} \end{bmatrix} = C^{-1} \begin{bmatrix} 1 \\ 0 \\ \\ 0 \end{bmatrix} = \begin{bmatrix} c'_{11} \\ c'_{21} \\ \\ c'_{n1} \end{bmatrix}$$

where the c'_{ij} are elements of C^{-1}.

Therefore, the constants k_{i1} corresponding to solution U_1 are determined to be as follows:

$$(7) \qquad \begin{cases} k_{11} = c'_{11}e^{m_1T} \\ k_{21} = c'_{21}e^{m_2T} \\ \\ k_{n1} = c'_{n1}e^{m_nT} \end{cases}$$

so that, substituting (7) back into (4) we get

$$(8) \qquad \begin{bmatrix} u_{11}(t) \\ u_{21}(t) \\ \\ u_{n1}(t) \end{bmatrix} = C \begin{bmatrix} k_{11}e^{-m_1t} \\ k_{21}e^{-m_2t} \\ \\ k_{n1}e^{-m_nt} \end{bmatrix} = C \begin{bmatrix} c'_{11}e^{m_1(T-t)} \\ c'_{21}e^{m_2(T-t)} \\ \\ c'_{n1}e^{m_n(T-t)} \end{bmatrix}$$

Therefore, if in general we choose the "special" solutions $U_j = [u_{1j}\ u_{2j} \ldots u_{nj}]^T$ to the adjoint system such that $u_{ij}(T) = \delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$

we get

$$(9)\qquad \begin{bmatrix} u_{1j}(t) \\ u_{2j}(t) \\ \\ u_{nj}(t) \end{bmatrix} = C \begin{bmatrix} c'_{1j}e^{m_1(T-t)} \\ c'_{2j}e^{m_2(T-t)} \\ \\ c'_{nj}e^{m_n(T-t)} \end{bmatrix}$$

so that, by grouping the n sets of solutions so obtained into a single matrix equation, we obtain

$$(10)\qquad \begin{bmatrix} u_{11} & u_{12} & & u_{1n} \\ u_{21} & u_{22} & & u_{2n} \\ & & & \\ u_{n1} & u_{n2} & & u_{nn} \end{bmatrix} = C \begin{bmatrix} e^{m_1(T-t)} & 0 & \cdots & 0 \\ 0 & e^{m_2(T-t)} & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & e^{m_n(T-t)} \end{bmatrix} C^{-1}.$$

Since the U_j are independent solutions to the adjoint system, a linear combination $U = U_1 + b_1U_2 + \ldots + b_{n-1}U_n$, where the b's are arbitrary constants, is also a solution. A solution obtained in this manner is called an adjoint vector.

If $\bar{e}_1, \bar{e}_2, \ldots \bar{e}_n$ represent the base vectors for an n-dimensional cartesian coordinate set, we can express the adjoint vector $\bar{U}$ as

$$(11)\qquad \bar{U} = \sum_i u_i \, \bar{e}_i$$

where

$$(12)\qquad u_i = u_{i1} + b_1u_{i2} + \ldots\ldots + b_{n-1}u_{in} \qquad (i = 1,2,\ldots,n).$$

For a minimum-time path we choose $\bar{F} = \bar{F}^+$ such that the components of $\bar{F}^+$ satisfy the relation

$$(13)\qquad f_i^+ = \operatorname{sgn} u_i \qquad\qquad (i = 1,2,\ldots,n).$$

From the Fundamental Adjoint Formula we have

$$(14)\qquad \left[\sum_i u_i x_i\right]_0^T = \int_0^T \left(\sum_i u_i f_i^+\right) dt$$

which, upon successive substitution of the "special" adjoint solutions U_j reduces to

$$(15)\qquad x_j(T) = \sum_i u_{ij}(0)x_i(0) + \int_0^T \left(\sum_i u_{ij} f_i^+\right) dt$$

for $j = 1,2,..,n$. Equation (15) can also be written in vector notation as

$$(16)\qquad x_j(T) = (\bar{U}_j.\bar{X})_0 + \int_0^T (\bar{U}_j.\bar{F}^+)\, dt \qquad (j = 1,2,..,n).$$

Again, as in the previous sections, we desire to drive the system from $X(0)$ to X_f in minimum time T. We therefore consider a linear combination of the $x_j(T)$, using the constants previously described in defining the adjoint vector. We have

$$(17)\qquad x_1(T) + b_1x_2(T) +\ldots+ b_{n-1}x_n(T) = (\bar{U}.\bar{X})_0 + \int_0^T (\bar{U}.\bar{F}^+)dt$$

where $\bar{U} = \bar{U}_1 + b_1\bar{U}_2 + \ldots + b_{n-1}\bar{U}_n$ is the adjoint vector.

Comparing the resulting terminal state $X(T)$ with the desired terminal state X_f, we define the component errors as

$$(18)\qquad \Delta X_j = X_{jf} - x_j(T) \qquad (j = 1,2,\ldots,n).$$

If any ΔX_j is not zero we must determine the changes in the input parameters required to effect X_f.

Suppose the initial values of the input parameters are replaced by nearby values

$$(19)\quad \begin{aligned} T &\longrightarrow T + \delta T \\ b_1 &\longrightarrow b_1 + \delta b_1 \\ &\vdots \\ b_{n-1} &\longrightarrow b_{n-1} + \delta b_{n-1}. \end{aligned}$$

Then the changes δX_j resulting in the terminal values $x_j(T)$ can be expressed as (for $j = 1,2,\ldots,n$)

$$(20)\quad \delta X_j = \left(\left(\frac{\partial \bar{U}_j}{\partial T}.\bar{X}\right)_o + \int_o^T \left(\frac{\partial \bar{U}_j}{\partial T}.\bar{F}^+\right)dt + f_j^+(T) - \sum_{i=1}^{n} \sum_{k=1}^{k_i} (u_{ij}[f_i^+])_{t_{ik}} \right) \delta T$$

$$+ \left[\sum_{i=1}^{n} \sum_{k=1}^{k_i} \left(\frac{u_{ij}[f_i^+]\, u_{i2}}{\dot{u}_i} \right)_{t_{ik}} \right] \delta b_1$$

$$\vdots$$

$$+ \left[\sum_{i=1}^{n} \sum_{k=1}^{k_i} \left(\frac{u_{ij}[f_i^+]\, u_{in}}{\dot{u}_i} \right)_{t_{ik}} \right] \delta b_{n-1}.$$

Equation (20) has the form of the total differential

$$(21)\quad \delta X_j = \frac{\partial X_j}{\partial T}\delta T + \frac{\partial X_j}{\partial b_1}\delta b_1 + \ldots\ldots + \frac{\partial X_j}{\partial b_{n-1}}\delta b_{n-1}$$

for $(j=1,2,\ldots,n)$.

Setting $\Delta X_j = \delta X_j$ and inverting the "correction" matrix defined by (21), we can determine the changes $\delta T, \delta b_1, \ldots, \delta b_{n-1}$ required to effect X_f, if the routine converges.

8. Conclusions and Comments.

The solution of the minimum-time problem for second and third-order systems was programmed and run on the CDC-1604 computer, using the Runge-Kutta-Gill method of numerical integration to calculate the required integrals. It has a desirable feature that it is easy to vary the time intervals any time during a run.

The convergence of the differential-correction routine appears to be fairly critical, especially for stable systems. That is, if for a given set of input parameters the error $X_f - X(T)$ is too large, the routine will not converge.

The switching times t_1, t_2 for second-order systems were determined exactly, so that the degree of convergence was limited only by the degree of precision provided in the numerical integration (limited by the size of Δt) and other routines. Convergence of the routine was obtained to an accuracy of better than six decimal places ($\Delta R^2 = 10^{-12}$).

For the third-order systems the switching times were not determined exactly: if a component of the optimum forcing function changed sign in a time interval Δt, it was assumed to have changed sign at the end-point of the interval. In this case, convergence of the routine was limited to about three decimal places ($\Delta R^2 = 10^{-6}$) for $\Delta t = .005$. Convergence to greater accuracy could be achieved by reducing the size of Δt in the numerical integration routine.

The method of solution appears to be applicable to more general problems. The case where the end-point X_f varies with time T may be handled by incorporating into the correction

equations (7.18) a term involving dX_i/dT, as was done by Faulkner and Ward [10] in a rocket trajectory problem.

The solution curves have several interesting properties. As noted in section 5, if the solution for a second-order system has two switching times, the difference is a constant of the system. A proof is given here that this is generally the case:

Let us consider a linear dynamic system represented by the system of differential equations

(1) $$\overset{\circ}{X} = A X + F,$$

which can be written in expanded form as

(2) $$\begin{cases} \overset{\circ}{x} = ax + by + f & |f| \leq 1 \\ \overset{\circ}{y} = cx + dy + g & |g| \leq 1 \end{cases}$$

where the elements a,b,c,d are constants, and the eigenvalues of A are real and distinct.

The eigenvalues of matrix A (roots of the characteristic equation $|A - mI| = 0$) are

(3) $$m_1 = \frac{a + d + r}{2}, \quad m_2 = \frac{a + d - r}{2}$$

where

(4) $$r = \sqrt{(a-d)^2 + 4bc} .$$

The "special" solutions U_1, U_2 (described in earlier sections) of the system adjoint to (1) can be expressed jointly in the following matrix equation:

(5) $$[U_1\ U_2] = \begin{bmatrix} u_1 & u_2 \\ v_1 & v_2 \end{bmatrix} = C \begin{bmatrix} e^{m_1(T-t)} & 0 \\ 0 & e^{m_2(T-t)} \end{bmatrix} C^{-1},$$

where C is the matrix of eigenvectors of $-A^T$.

The general solution $U = U_1 + kU_2$ of the adjoint system can then be expressed as

$$(6) \quad U = \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} u_1 & u_2 \\ v_1 & v_2 \end{bmatrix} \begin{bmatrix} 1 \\ k \end{bmatrix} = \left(C \begin{bmatrix} e^{m_1(T-t)} & 0 \\ 0 & e^{m_2(T-t)} \end{bmatrix} C^{-1} \right) \begin{bmatrix} 1 \\ k \end{bmatrix}$$

Let t_1, t_2 denote the switching times of f and g respectively, i.e., $u(t_1) = v(t_2) = 0$.

Solving for C and C^{-1} by the methods described in Appendices I and II, we obtain the following expression for u and v evaluated at t_1 and t_2 respectively:

$$(7) \quad \begin{cases} u(t_1) = 0 = e^{m_1(T-t_1)} \left[\dfrac{a-d+r+2kc}{2r} \right] + e^{m_2(T-t_2)} \left[\dfrac{-a+d+r-2kc}{2r} \right] \\[2ex] v(t_2) = 0 = e^{m_1(T-t_1)} \left[\dfrac{2b+k(-a+d+r)}{2r} \right] + e^{m_2(T-t_2)} \left[\dfrac{-2b+k(a-d+r)}{2r} \right] \end{cases}$$

A simplification of equations (7) can be made by using the relation $r = m_1 - m_2$. The reduced form is then

$$(8) \quad \begin{cases} e^{r(T-t_1)} = \dfrac{a - d + 2kc - r}{a - d + 2kc + r} \\[2ex] e^{r(T-t_2)} = \dfrac{2b - kr - k(a-d)}{2b + kr - k(a-d)} \end{cases}$$

If we now consider the ratio

$$(9) \quad \frac{e^{r(T-t_1)}}{e^{r(T-t_2)}} = e^{r(t_2-t_1)} = \frac{a-d+2kc-r}{a-d+2kc+r} \cdot \frac{2b+kr-k(a-d)}{2b-kr-k(a-d)}$$

which can be reduced to

$$(10) \quad e^{r(t_2-t_1)} = \frac{2\ (a - d - r)}{2\ (a - d + r)} \cdot \frac{b - k(a-d) - k^2c}{b - k(a-d) - k^2c} ,$$

we obtain the following relation between t_1 and t_2:

$$(11) \quad t_2 - t_1 = \frac{1}{r} \log_e \left| \frac{a - d - r}{a - d + r} \right| ,$$

that is, the difference between the switching times t_1 and t_2 on a minimum-time trajectory is an invariant of the dynamic system.

On the other hand, switching curves or surfaces, which have received much attention, seem to have no particular significance. They are not implicit in the differential equations, but depend on the particular problem. Typical switching curves are given in Illustration 2 for the case where the trajectory begins at the origin. In the same figure, the dotted lines indicate the boundaries to the points of static stability.

The concept of using the adjoint system to calculate differentials of end values is apparently due to Bliss in his work in ballistics [8]. In the problems solved here, an explicit solution to the adjoint system was obtained, which made possible the integral formulas for the end-point X(T).

BIBLIOGRAPHY

1. R. Bellman, et al., "On the Bang-Bang Control Problem" Qtr.Appl.Math.,14, Apr 1956, pp 11-18.

2. C.A.DeSoer, "The Bang-Bang Servo Problem Treated by Variational Techniques," Information and Control 2, pp 333-348, 1959.

3. J.P. LaSalle, "The Time Optimal Control Problem" In Contributions to the Theory of Nonlinear Oscillations, Vol V, pp 1-24, Princeton Univ. Press, 1960.

4. F.D. Faulkner, Lecture Notes on the Adjoint System of Differential Equations.

5. W.Kaplan, Advanced Calculus, Addison-Wesley, 1952.

6. R. Bellman, Stability Theory of Differential Equations, McGraw-Hill, 1953.

7. L.S. Pontryagin, "Optimal Control Processes" in "Control Systems and Data Handling"

8. G.A. Bliss, The Use of Adjoint Systems in the Problem of Differential Corrections for Trajectories" J. of the U.S. Artillery, LI, pp 296-311, 1919.

9. G.A. Bliss, Mathematics for Exterior Ballistics, Wiley, 1944.

10. F.D. Faulkner, E.N. Ward, "Optimum Interception of a Ballistic Missile at Moderate Range", United States Naval Postgraduate School, 1961.

11. Coddington and Levinson, Theory of Ordinary Differential Equations, McGraw-Hill, 1955.

APPENDIX I

Solution of A System of Homogeneous Linear First-Order Differential Equations with Constant Coefficients.

Suppose we have the differential equation

(1) $\quad \overset{\circ}{U} = B\,U$

where B is an n x n matrix with constant elements b_{ik} and with real, distinct eigenvalues.

Let us assume a solution of (1) of the form

(2) $\quad U = e^{mt}\,C$

where C is an n x 1 column matrix with constant elements. Differentiating (2) with respect to t, we get

(3) $\quad \overset{\circ}{U} = m\,e^{mt}\,C.$

Substitution of (3) and (2) into (1) gives the following result:

(4) $\quad m\,e^{mt}\,C = B\,e^{mt}\,C$

and, since e^{mt} is a scalar quantity appearing on both sides of (4), we have

(5) $\quad m\,C = B\,C.$

Therefore, if m is an eigenvalue of matrix B, then C must be a corresponding eigenvector. Since there exist n distinct eigenvalues $m_1, m_2, \ldots, m_n$ of matrix B, there are n independent solutions of (1) of the form[1]

(6) $\quad U^j = e^{m_j t}\,C^j \qquad (j = 1,2,\ldots,n)$

where $C^j = [c_{1j}\ c_{2j}\ \ldots\ c_{nj}]^T$ is the eigenvector corresponding to eigenvalue m_j, and $U^j = [u_1^j\ u_2^j\ \ldots\ u_n^j\ ]^T.$

1. In this section only, m_j is associated with solutions for U.

The general solution of (1) is then a linear combination of the U^j, i.e.,

$$U = \sum_j k_j U^j = \sum_j k_j e^{m_j t} C^j \tag{7}$$

where the k_j are arbitrary constants.

Componentwise, (7) may be written as

$$u_i = \sum_j k_j e^{m_j t} c_{ij} \qquad (i = 1,2,\ldots,n). \tag{8}$$

APPENDIX II

Determination of the Components of Eigenvector C^j.

We have the adjoint system of differential equations

(1) $\dot{U} = -A^T U$

where A is an n x n matrix with constant elements a_{ik} and with real, distinct eigenvalues.

From Appendix I we know that the solution of (1) is of the form

(2) $U = \sum_j k_j e^{-m_j t} C^j$

where the $-m_j$ are the eigenvalues of matrix $-A^T$ and the C^j are the corresponding eigenvectors of $-A^T$. The k_j are arbitrary constants.

Differentiating (2) with respect to t we get

(3) $\dot{U} = - \sum_j m_j k_j e^{-m_j t} C^j$

The i^{th} component of (3) can be written as

(4) $\dot{u}_i = - \sum_j m_j k_j e^{-m_j t} c_{ij}$, $(i = 1,2,\ldots,n)$.

We can also write the i^{th} component of (1) as

(5) $\dot{u}_i = - a_{1i}u_1 - a_{2i}u_2 - \ldots\ldots - a_{ni}u_n$, $(i=1,2,\ldots,n)$.

The i^{th} component of (2) can likewise be written as

(6) $u_i = \sum_j k_j e^{-m_j t} c_{ij}$, $(i = 1,2,\ldots,n)$.

Substituting (6) into (5) we get

(7) $\dot{u}_i = -a_{1i}(\sum_j k_j e^{-m_j t} c_{1j}) - a_{2i}(\sum_j k_j e^{-m_j t} c_{2j}) - \ldots.$

Since (4) and (7) represent equivalent expressions of the $\dot{u}_i$ we can equate the corresponding coefficients of the $k_j e^{-m_j t}$. For example, equating the corresponding coefficients of $k_1 e^{-m_1 t}$ we get

(8) $-m_1 c_{i1} = -a_{1i} c_{11} - a_{2i} c_{21} - \dots - a_{ni} c_{n1}$, $(i = 1,2,\dots n)$.

Therefore, writing out all n of the equations obtained by equating corresponding coefficients of $k_1 e^{-m_1 t}$ we get

$$
(9) \quad \begin{aligned}
-m_1 c_{11} &= -a_{11} c_{11} - a_{21} c_{21} - \dots\dots - a_{n1} c_{n1} \\
-m_1 c_{21} &= -a_{12} c_{11} - a_{22} c_{21} - \dots\dots - a_{n2} c_{n1} \\
\\
-m_1 c_{n1} &= -a_{1n} c_{11} - a_{2n} c_{21} - \dots\dots - a_{nn} c_{n1}
\end{aligned}
$$

which can be written as the following matrix equation:

$$
(10) \quad \begin{bmatrix} a_{11}-m_1 & a_{21} & & a_{n1} \\ a_{12} & a_{22}-m_1 & & a_{n2} \\ & & & \\ a_{1n} & a_{2n} & & a_{nn}-m_1 \end{bmatrix} \begin{bmatrix} c_{11} \\ c_{21} \\ \\ c_{n1} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \\ 0 \end{bmatrix}
$$

Denote by Δ the determinant of coefficients in (10). For a nontrivial solution of the c_{i1} to exist it is necessary that $\Delta = 0$, i.e., that

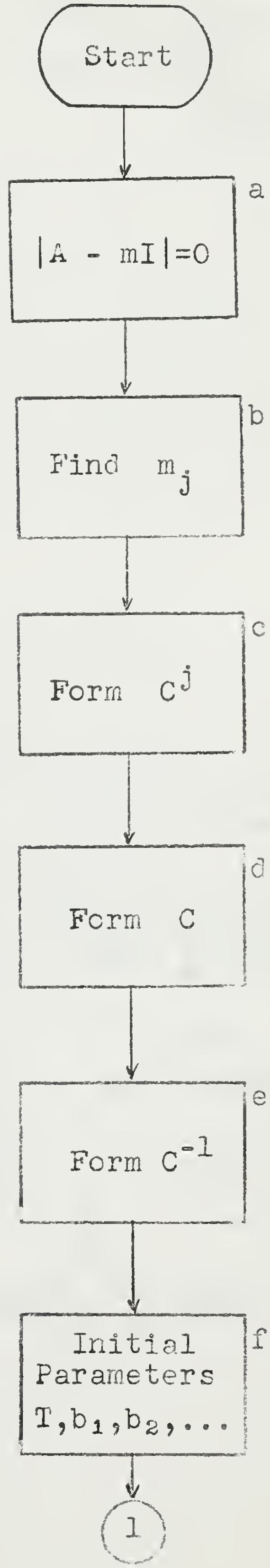
(11) $|A^T - m_1 I| = 0$.

A solution of the c_{i1} consists of the cofactors of the elements of a row of the matrix $[A^T - m_1 I]$, see Bellman [6]. This is the same as the cofactors of a column of $[A - m_1 I]$. That is, the components of eigenvector C^1 are in the same ratio as the cofactors of the elements of a column of the characteristic matrix $[A - m_1 I]$.

Similarly, the components of the remaining eigenvectors C^j can be found by using successively the cofactors of the elements of a column of the matrices $[A - m_j I]$.

Illustration 1

Computer Flow Chart for Solution of N^{th} Order Minimum-Time Problem



a. Form Characteristic Equation of matrix A.

b. Find eigenvalues m_j of A (roots of Characteristic Eq.).

c. Form eigenvectors C^j of $-A^T$.
$C^j = [c_{1j}\ c_{2j}\ \ldots\ c_{nj}]^T$

d. Form C, the matrix of eigenvectors of $-A^T$.

e. Invert C to get C^{-1}.

f. Load initial values of the input parameters.

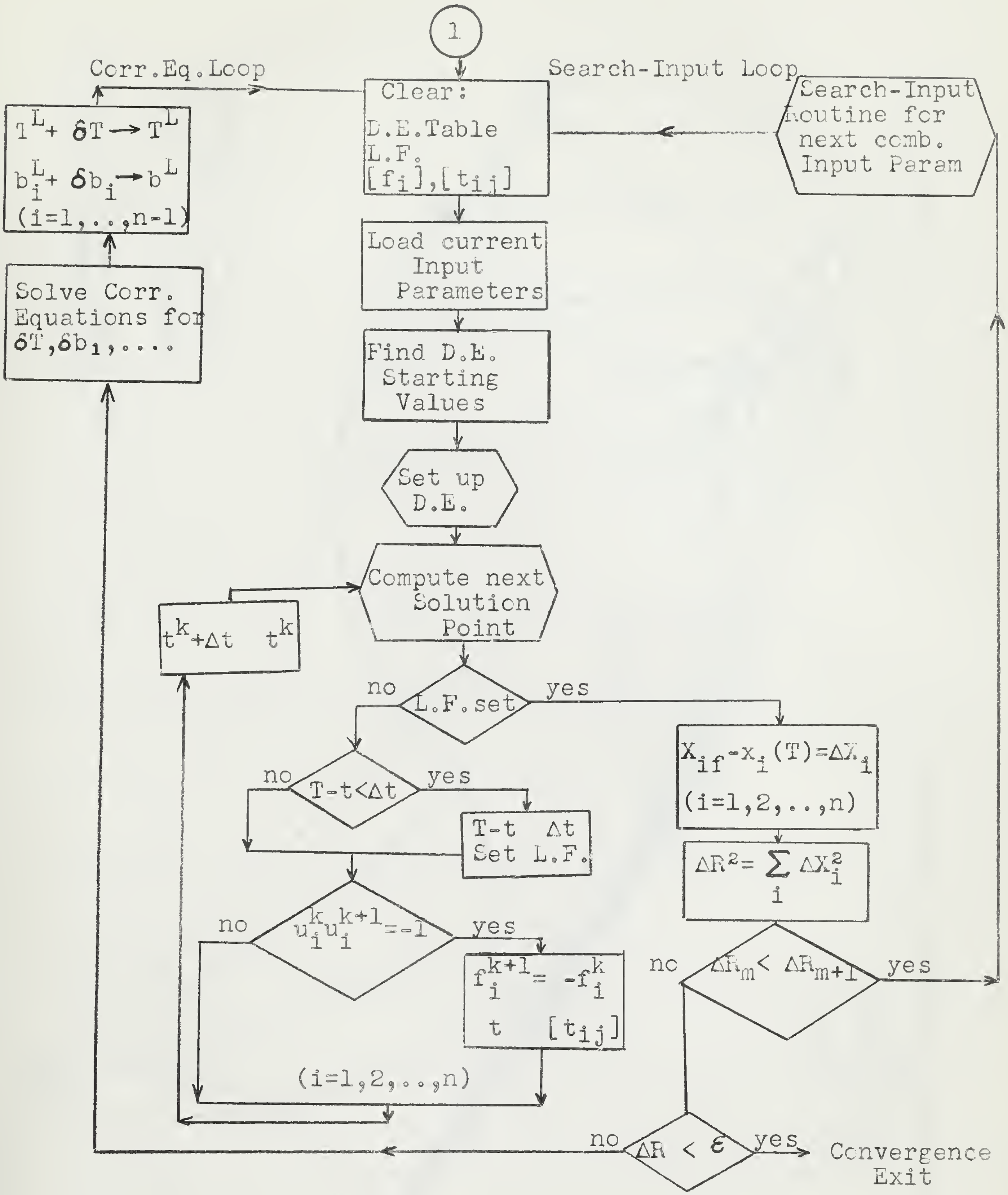
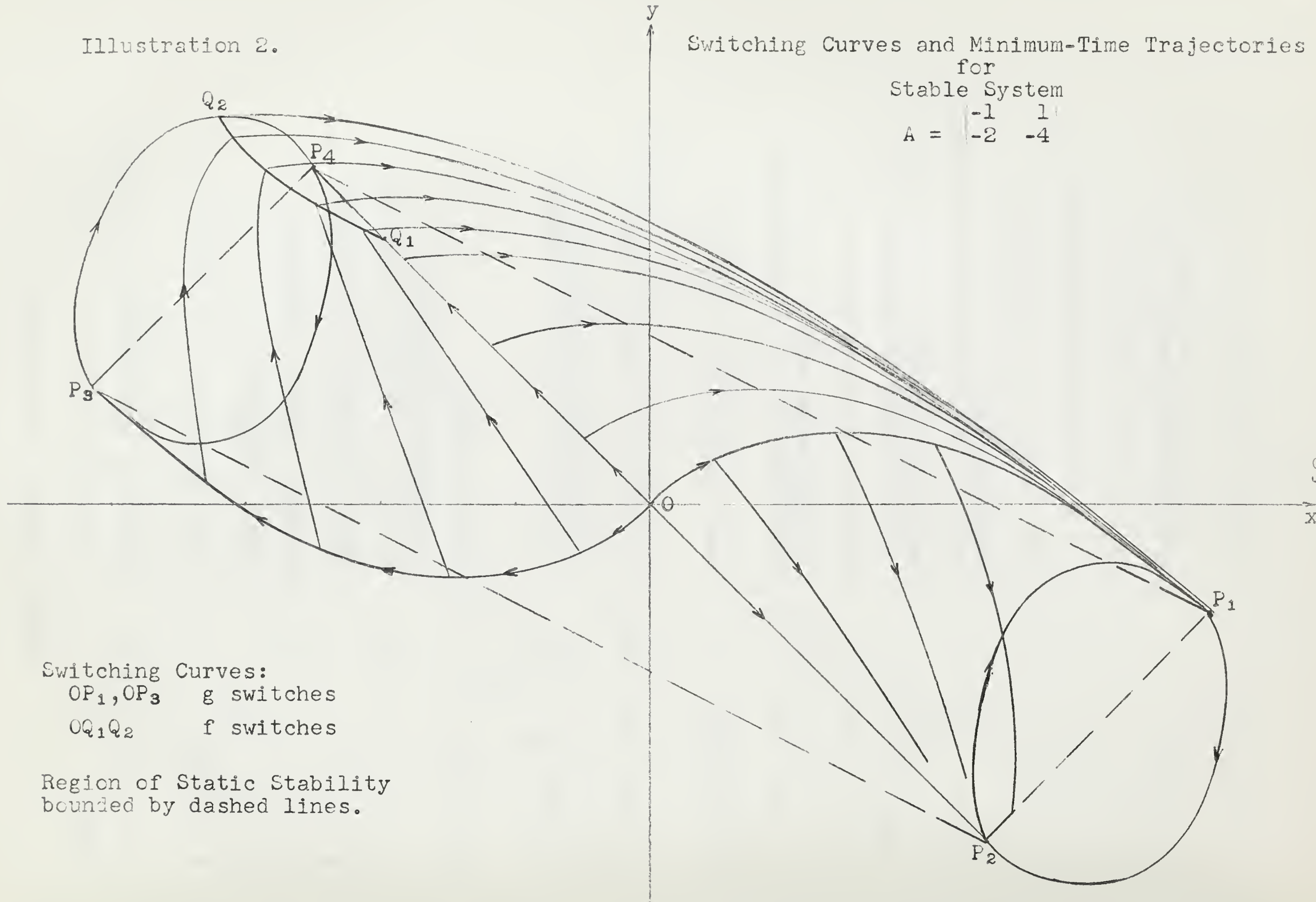

1
Corr.Eq.Loop
Search-Input Loop
Clear:
D.E.Table
L.F.
[fi],[tij]
Search-Input
Routine for
next comb.
Input Param
T^L + δT → T^L
b_i^L + δb_i → b^L
(i=1,...,n-1)
Load current
Input
Parameters
Solve Corr.
Equations for
δT,δb1,....
Find D.E.
Starting
Values
Set up
D.E.
Compute next
Solution
Point
t^k+Δt t^k
no
L.F.set
yes
Xif-xi(T)=ΔXi
(i=1,2,..,n)
no
T-t<Δt
yes
T-t Δt
Set L.F.
ΔR²= Σ ΔXi²
i
no
ui^k ui^k+1 =-1
yes
fi^k+1 = -fi^k
t [tij]
no
ΔRm < ΔRm+1
yes
(i=1,2,...,n)
no
ΔR < ε
yes
Convergence
Exit

Illustration 2.

Switching Curves and Minimum-Time Trajectories
for
Stable System

$$A = \begin{vmatrix} -1 & 1 \\ -2 & -4 \end{vmatrix}$$



Switching Curves:
OP_1, OP_3 g switches
OQ_1Q_2 f switches

Region of Static Stability bounded by dashed lines.

SPECIAL SYMBOLS AND NOTATION

Symbol	Meaning
$X = [x_1\ x_2\ ..\ x_n]^T$	Matrix of coordinates, also called System state function.
$\bar{X} = \sum_i x_i \bar{e}_i$	Same, considered as a vector.
$\bar{R} = x\bar{e}_1 + y\bar{e}_2$	Same, considered as a vector.
$F = [f_1\ f_2\ ..\ f_n]^T$	Forcing (Control) function.
$\bar{F} = \sum_i f_i \bar{e}_i$	Same, considered as a vector.
$U = [u_1\ u_2\ ..\ u_n]^T$	Solution to adjoint system of differential equations.
$\bar{U} = \sum_i u_i \bar{e}_i = \sum_j b_{j-1} U_j$ $(b_0 = 1)$	Adjoint vector
$U^j = [u_1^j\ u_2^j\ ..\ u_n^j]^T$	j^{th} solution of adjoint system corresponding to eigenvalue $-m_j$.
$U_j = [u_{1j}\ u_{2j}\ ..u_{nj}]^T$	j^{th} "special" adjoint solution such that $u_{ij}(T) = \delta_{ij} = \begin{cases} 1, & i=j \\ 0, & i \neq j \end{cases}$
$\bar{U}_j = \sum_i u_{ij} \bar{e}_i$	Same, considered as a vector.
$C^j = [c_{1j}\ c_{2j}\ ..c_{nj}]^T$	Eigenvector of $-A^T$ corresponding to eigenvalue $-m_j$.
$C = [C^1\ C^2 ...\ C^n]$	Matrix of eigenvectors of $-A^T$.
$(\bar{U}.\bar{F}) = \sum_i u_i f_i$	Scalar product of adjoint vector and forcing vector.
$\bar{e}_1, \bar{e}_2, ..., \bar{e}_n$	Base vectors of Cartesian coordinate set in n-space.
$A = (a_{ik})$	n x n matrix of coefficients.
$B = (b_{ik})$	n x n matrix of coefficients.
$[A - mI]$	Characteristic matrix of A.
$[A - m_jI]$	Characteristic matrix of A associated with eigenvalue m_j.
$\|A - mI\| = 0$	Characteristic equation of A.

Symbol	Meaning
E	Diagonal matrix with j^{th} diagonal element $e^{m_j(T-t)}$
$X(0) = (x_1, x_2, ., x_n)_o$	Initial state of dynamic system
$X_f = (X_{1f}, X_{2f}, .., X_{nf})$	Terminal state of dynamic system (specified)
$X(T) = (x_1, x_2, .., x_n)_T$	Terminal state of dynamic system at time $t = T$.
T	Terminal value of time; also transpose when used with matrix.
f	Subscript f denotes final value.
$-$	Bar over a symbol indicates that it is a vector.
c'_{ij}	Element of C^{-1}.
$\Delta X_i = X_{if} - x_i(T)$	Error in terminal value of i^{th} component.
$I_k = (\bar{U}.\bar{F}^+)$, $t_k < t < t_{k+1}$	
$\delta X_i = \frac{\partial X_i}{\partial T}\delta T + \frac{\partial X_i}{\partial b_1}\delta b_1 + \dots + \frac{\partial X_i}{\partial b_{n-1}}\delta b_{n-1}$	
$[\quad]_{t_{ik}}$	jump discontinuity in variable at time t_{ik}
$[f_i]_{t_{ik}} = f_i(t_{ik}+0) - f_i(t_{ik}-0)$	
$(\bar{U}_j.[\bar{F}])_{t_{ik}} = (\bar{U}_j.\bar{F})_{t_{ik}+0} - (\bar{U}_j.\bar{F})_{t_{ik}-0} = \left(u_{ij}[f_i]\right)_{t_{ik}}$	
t_{ik}	k^{th} switching time of f_i.
$K = [k_1 e^{-m_1 t} \;\; k_2 e^{-m_2 t} \;\dots\; k_n e^{-m_n t}]^T$	